DOCUMENT RESUME

ED 185 767 EC 123 922

AUTHOR Jenkins, Joseph R.; And Others
TITLE Measuring Pupil Progress Toward the Least Restrictive Environment.
INSTITUTION Minnesota Univ., Minneapolis. Inst. for Research on Learning Disabilities.
SPONS AGENCY Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
REPORT NO IRLD-Mono-10
PUB DATE Aug 79
CONTRACT 300-77-0491
NOTE 32p.; See also EC 123 901-925.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Accountability; Compliance (Legal); Data Collection; *Disabilities; Elementary Secondary Education; *Evaluation Methods; Federal Legislation; *Mainstreaming; *Student Evaluation; Student Placement
IDENTIFIERS Education for All Handicapped Children Act

ABSTRACT

The authors consider reasons for increased emphasis on measuring pupil progress in special education, the use made of pupil progress data, the desirable qualities of measurement systems, and some problems associated with current commercially available data systems. Techniques for measurement are discussed, based on the requirements of P.L. 94-142, the Education for All Handicapped Children Act, for the least restrictive environment (LRE). The authors assert that measurement rooted in the principle of LRE is more consistent with the intents of the law and more useful. In addition, they suggest that such measurement better meets the requirements for technical adequacy. (Author/CL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED185767

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James Ysseldyke

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IRLD

Director: James E. Ysseldyke

Associate Director: Phyllis K. Mirkin

The Institute for Research on Learning Disabilities is supported by a contract (300-77-0491) with the Bureau of Education for the Handicapped, Department of Health, Education, and Welfare, U.S. Office of Education, through Title VI-G of Public Law 91-230. Institute investigators are conducting research on the assessment/decision-making/intervention process as it relates to learning disabled children. Research activities are organized into eight major areas:

I. Adequacy of Norm-Referenced Data for Prediction of Success

II. Computer Simulation Research on the Assessment/Decision-making/Intervention Process

III. Comparative Research on Children Labeled LD and Children Failing Academically but not Labeled LD

IV. Surveys on In-the-Field Assessment, Decision Making, and Intervention

V. Ethological Research on Placement Team Decision Making

VI. Bias Following Assessment

VII. Reliability and Validity of Formative Evaluation Procedures

VIII. Data-Utilization Systems in Instructional Programming

Additional information on these research areas may be obtained by writing to the Editor at the Institute.

The materials presented herein were prepared under government sponsorship. Contractors are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position of the Bureau of Education for the Handicapped.

Monograph No. 10

MEASURING PUPIL PROGRESS TOWARD THE LEAST RESTRICTIVE ENVIRONMENT

Joseph R. Jenkins

University of Washington

Stanley L. Deno and Phyllis K. Mirkin

Institute for Research on Learning Disabilities

University of Minnesota

August, 1979

Measuring Pupil Progress Toward the Least Restrictive Alternative

Although we have been in the business of developing educational programs for special education students for many years now we have had very few urgent discussions of how to measure student progress in achieving the goals and objectives established in these programs. Lately, however, this topic has begun to appear and reappear with great frequency. Why the change? Several events appear crucial.

Before 1975, special education programs were relatively secure from scrutiny, operating in the safety of isolation and privacy. Parents of handicapped children occasionally registered a complaint against the system, but the system was usually robust enough and sufficiently opaque to withstand any serious attempt to understand or modify it. While there was occasional embarrassment such as the efficacy research (i.e., Goldstein, Moss, & Jordan, 1965) which suggested that despite expensive programs, specially trained teachers, unique grouping arrangements, reduced class sizes, and special curriculum modifications, it could not be shown that handicapped children made any better educational progress than if they were given "non-special" education. This research along with occasional complaints and inquiries from parents and regular educators produced only minor annoyances for special education. It did not require administrators to produce any systems-wide change, and though some administrators attempted to establish more accountability, they were few in number and did so voluntarily.

FEB 1 3 1980

Since 1975 and the passage of P.L. 94-142, matters have changed dramatically. The requirement that to the maximum extent possible handicapped children be educated in programs that include non-handicapped children -- that is, in educational environments that are as close as possible to normal, as well as the requirement that Individual Educational Programs (IEPs) be developed for every handicapped student, has had an enormous impact on the current status of special education. The IEPs must include goals and objectives related to the child's instructional and therapeutic needs and in some districts must be accompanied by a time-line so that progress toward their attainment can be monitored during the course of an intervention. Parents and guardians must be permitted to participate in the decision which determines the Least Restrictive Environment (LRE) and must be given both sufficient information to understand the plans for their children and an opportunity to exercise their right to approve or reject these plans.

It is obvious how the legal requirements for IEPs and parent involvement would raise our sensitivities about the need to develop satisfactory procedures for measuring children's progress. Programs, schools, and teachers must specify what they intend to accomplish with each child, measure the degree to which their intended accomplishments are realized, and, finally, make this information available to concerned parties (parents, other professionals, state departments, etc.). Although it may be less immediately obvious how the legal requirement for LRE relates to the issue of measurement, the relationship is actually very direct, particularly in programs for mildly handicapped students. The LRE requirement implies

that handicapped children must be taught skills that are most likely to result in placement or maintenance in regular classrooms. To ensure that this is being done, handicapped children's progress must be quantified in relation to progress on curriculum objectives established by and for regular education. The measurement procedure employed must depict handicapped students' progress in relation to the minimal requirements for placement in less restrictive environments Thus, in accord with this line of reasoning, the Least Restrictive Environment has much to say about what will be measured and how often measurement must occur.

Uses of Pupil Progress Measurement

What are the specific uses of a measurement system in relation to this overall goal of providing appropriate educational programs to children in least restrictive environments?

Eligibility Decisions

One use which comes immediately to mind is to determine eligibility. A measurement system is needed to determine where a child stands relative to his peers in mastering mainstream tasks. Such information is necessary to those who must determine whether a child qualifies for special education services.

Program Planning Decisions

A second area in which a measurement system is useful is in determining the type of educational program to be provided. Ideally, of course, the same measurement system which has been used to determine eligibility should also serve to make program planning decisions. Unfortunately, however, a common concern of educational decision makers is the lack of congruence

between the data used for diagnosis and that used for program planning. Researchers at the Institute for Research on Teaching (1979), for example, report that when reading clinicians diagnosed and proposed remediation for children with reading problems, the diagnosis and proposed remediation were often unrelated or the diagnoses "appeared too vague to guide treatment."

A measurement system which uses the same data for initial diagnosis and program planning would help to overcome this hiatus between diagnosis and treatment.

Program Adjustment Decisions

A third major use of a data system is in making adjustments in programs during implementation. A measurement system is needed which allows the program implementator, usually the teacher, to stay close enough to the child for effective decision making. These procedures are not part of the usual classroom practice (Goodlad & Klein, 1974; Mirkin, 1978; Peterson, Marx, & Clark, 1978; Popham & Baker, 1970; Zoharik, 1975).

Program adjustments might include, among others, changes in instruction (e.g., modeling to flash card drill), changes in incentives (e.g., tangible incentives to points), changes in curriculum (e.g., from Distar to Scott Foresman), as well as changes in the level and frequency of special education services.

The purpose of these adjustments is, of course, to increase the progress of the students toward achievement of established goals and objectives. The assumption is that the chances of solving a problem successfully are increased if multiple rather than single solutions are proposed and implemented. In such an approach, measurement provides the data necessary to test the effectiveness of different program changes

and to determine which of those program changes should be maintained.

Program Effectiveness Decisions

The final use of a measurement system in the development of educational programs for children in the least restrictive environment is for appraising overall program success. A measurement system should be capable of providing information concerning the extent to which special education services have been successful in improving pupil progress and maintaining children in less restrictive placements. Although other specific uses for a measurement system could be identified, and the uses that are listed here can be reorganized, the decision areas listed would appear to be central to most special education programs.

Desirable Characteristics

What are some desirable characteristics of a data system for each of these uses?

Relevant

One characteristic is that the system generates data that provide a fair representation of what is being taught and how well the student is progressing on the tasks required in the less restrictive environment. This information is critical for eligibility and planning decisions as well as program adjustment and effectiveness decisions.

Sensitive

Another important characteristic is that the data generated be sensitive to real changes in performance. Such sensitivity is essential if professionals intend to use data for making day-to-day or week-to-week adjustments in the child's program.

Flexible

A related characteristic is that the data system be flexible and adaptive enough to different instructional objectives. For example, if accurate oral reading is an instructional objective, then the data system should reflect directly on this objective, and likewise for other types of objectives, e.g., phonics, comprehension, and workbook exercises. Such flexibility is necessary if a data system is to make sense to its users.

Repeatedly Administrable

Another desirable characteristic of the data system is that it generate information with sufficient frequency for the particular decision areas (e.g., placement, program adjustments, overall program evaluation). For example, overall program evaluation and placement decisions do not require frequent measurement. In contrast, program adjustments related to instructional, incentive and curricular decisions, resource utilizational decisions, and least restrictive changes may require regularly and frequently generated data.

Easily Administrable

A final desirable characteristic of a measurement system is ease of administration. If data are to be collected frequently, the system cannot be time consuming or costly to implement.

Adequacy of Current Data Systems

How well do data systems currently available to educators satisfy these requirements? Of all the different types of data systems used in special education, pre-posttest administration of commercially available achievement tests probably enjoys the widest use, presumably because of

the tests' face validity and because of what they seem to promise the consumer. To determine how well standardized achievement tests satisfy the requirements of a good data system, let us examine some research that has been conducted on these instruments.

Jenkins and Pany (1978) compared the overlap between the content of a number of popular basal reading series and a number of standardized achievement tests commonly used in assessment of handicapped students. Hypothetical grade equivalent scores were computed for each test based on the assumption that the student had learned to read all the words taught in a particular reading series. Table 1 shows the grade equivalence produced by this procedure.

- -

Insert Table 1 about here

- -

As can be seen, some achievement tests appear to be more sensitive than others in measuring what is taught in a particular curriculum. This differential sensitivity seriously challenges the practice of using achievement tests for determining special education eligibility. Suppose, for example, that in a particular district a child must score at 50% or more below grade level in order to qualify for special services. The student's eligibility for special services could well be determined by the extent to which the achievement test administered did or did not include the content of the reading series used for instruction. Hypothetically, a beginning third grade student who had been taught in the Keys to Reading series would score more than 50% below grade level on each of the MAT subtests, but at or above grade level on the PIAT, SORT, and WRAT.

Similar problems would be expected were achievement test grade equivalent scores used to make placement decisions in a particular reading series. If the hypothetical grade equivalent scores developed by Jenkins and Pany are at all accurate, a student who mastered first grade Reading 360 vocabulary would obtain a SORT grade equivalent of of 1.4 whereas if the same student had read in Keys to Reading, a score of 2.2 could be expected. What is the proper interpretation of these results? Do they indicate that a student using Reading 360 is not really prepared to read second grade materials, but that one using Keys to Reading is? What is the correct placement for this child? It would certainly seem inappropriate to make this decision on the basis of achievement test scores.

Achievement tests also lack utility for purposes of program adjustment as they cannot be given with any regularity without invalidating the test. To determine the effectiveness of a particular instructional procedure, incentive, or curriculum for a particular child, data must be available at frequent and in some instances daily intervals.

Using achievement tests for program evaluation would also appear to be quite hazardous, yet students, teachers, and curricula are all subject to evaluation based on standardized test scores. For a particular student, the scores are often used to measure the amount of growth over some period. Children making "normal" progress are expected to advance one full grade equivalent for each year spent in school. The scores shown in Table 1, however, reveal that for any curriculum, the amount of growth varies with the particular test employed. Hypothetically, a child who had learned the content words in Grade 1 of Keys to Reading would, by the end of Grade 2,

gain one year and three months according to the PIAT, gain four months in word knowledge (MAT), lose two months in word analysis (MAT), but gain eight months according to the SORT and the WRAT.

A study by Armbruster, Stevens, and Rosenshine (1977) demonstrates that such problems are not confined to achievement tests that focus on measures of word recognition. These investigators categorized the content of two reading achievement tests -- the Metropolitan and California Achievement Test, along with a portion of the exercises contained in three reading series (Economy, Ginn 360, and Houghton-Mifflin). They concluded that the test items on the MAT or the CAT fail to cover roughly 64% of Economy, 65% of Ginn and 79% of Houghton-Mifflin's reading comprehension exercises. Moreover, there were rather large differences in relative emphasis between the reading series and the achievement tests. Based on the percent of different types of reading comprehension exercises and the percent of items tapping these different types of reading comprehension skills, the Ginn series correlated only .10 with the MAT. In contrast, a correlation of .43 was found between the CAT and Houghton-Mifflin. The implication of these data is that where the contents of a curriculum and an achievement test are poorly correlated, the test will be insensitive to gains produced by the curricula. The validity of the analyses by Jenkins and Pany (1978) and Armbruster et al. (1977) is corroborated in a study by Eaton and Lovitt (1972) based on actual student performance. They reported enormous inconsistency in the scores obtained by the same learning disabled students on different achievement tests and in the degree to which these achievement tests reflected the students' actual performance in particular curricula.

Another problem that arises when standardized achievement tests are used in program evaluation concerns the use of normative data. Most achievement tests provide only once a year, or at best twice a year, normative data (e.g., fall and spring). Unfortunately, referrals to special education do not always coincide with these testing dates. If a child is tested at a time that does not coincide with the norming dates, then the norms cannot be applied to that child whether these norms are reported in terms of grade equivalent scores, percentiles, or normalized standard scores. Even when students are tested during periods for which norming data are available, there are several problems with the typical scores generated from achievement tests.

One problem is the use of grade equivalent scores (GESs). GESs can never be used for program evaluations as they are not equal interval units and thus cannot be aggregated and averaged. Publishers do not really test children who are more than one grade level below the level of a particular test. For example, third grade children are not tested on a test for fifth grade level children. Thus GESs are only estimates and do not represent scores actually achieved by students who score more than one year below grade level. This, of course, is a crucial problem for handicapped children who almost always score several levels below their expected grade level. Finally, publishers produce grade equivalent norms in such a way that handicapped students could be shown to make more than one month's growth per month in school between testing periods; yet, in fact, these children will actually be falling farther and farther behind their peers (Tallmadge, 1977). Finally, GESs give misleading information with respect to students' relative standings. For example, students scoring one grade level below normal could, in fact, be in the 5th percentile,

10th percentile, 25th percentile, or even the 40th percentile, depending upon their actual grade level and on the particular achievement test administered.

Not all achievement test scores are as misleading as GESs. Percentiles, for example, provide accurate information on relative standing, but even these cannot be aggregated and averaged, since they do not represent equal interval units. Most problems can be solved by using normal curve equivalences (NCEs) that were developed for use with Title 1 evaluations (Tallmadge, 1976). Even NCEs are probably not satisfactory for special education program evaluations, however, because their use requires that achievement test administration coincide with norming periods. This, as we have seen, is a problem in special education since referral to programs is a process that occurs throughout the year, not just during those periods when tests were normed.

In summary, standardized achievement tests have definite liabilities for uses in determining special education eligibility, making placements in curricula, assessing the effects of program adjustments, and in overall program evaluation.

Measurement in the Least Restrictive Environment

One solution to the problems encountered when using commercially available tests that have been developed external to specific school programs is to train teachers to develop or use data systems which are rooted in the least restrictive environment. The approach is deceptively simple and straightforward. In such systems teachers take IEP objectives and use direct observation and recording of performance on those objectives over time to monitor pupil progress. Two alternative approaches to

measurement in the least restrictive environment have been identified (Deno & Mirkin, 1977).

Daily Performance Measurement

Daily performance measurement is designed to describe how a student's behavior changes on a single task over time. The measurement techniques of Precision Teaching thoroughly outlined in White and Haring (1976) clearly illustrate performance measurement. The salient characteristics of this approach are direct, daily measurement of student performance on what is being taught, and graphic data display.

The measurement in this system is considered direct in the sense that it focuses directly on what is taught. For example, if a student is being taught to say sounds for letters of the alphabet, then it is that behavior which is measured directly. If a student is being taught to write addition facts, then that behavior is measured. If a student is being taught to read orally and comprehend factual information, then oral reading proficiency and answers to factual comprehension questions are the focus of the measurement.

The use of daily performance measurement is depicted in Figure 1. In this case the teacher devised procedures to record the oral reading correct responses of a student and used these data to evaluate a particular instructional intervention, flashcard drill. As is evident, the data suggest that, for this student, flashcard drill favorably influenced the student's reading in context.

- - - - - - - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - - - - - - -

In terms of the desirable characteristics of a data system, daily performance measurement has much to offer. First, a teacher can be assured that the data are relevant to what is taught, since the measurement focuses on just that. Additionally, the data should be quite sensitive to changes in student performance that occur as a function of teaching. Data are gathered with sufficient frequency to evaluate program adjustments, e.g., instructional changes or incentive changes. Too, the data system is quite adjustable to any performance that is taught; for example, oral reading, writing spelling words, answering comprehension questions, saying math facts, and completing work book exercises.

How useful is this system for other decision areas such as eligibility determination, placement decisions, program adjustments, and summative evaluations? Although not invented as an eligibility tool, daily performance measurement can be adapted for determining a child's eligibility for special education. Specifically, these adaptations require that one identify a performance area that is important (e.g., answering comprehension questions), determine minimal acceptable performance in this area, and then compare a particular student's performance in this area with that of the standard. Deno and Mirkin (1977) have proposed such a procedure for determining eligibility of children who deviate from their peers in terms of social behavior, e.g., creating noise, physical contact, off-task, etc. The frequency of the target child's emission of these behaviors is observed over several days. Also recorded is the frequency of these behaviors by the target child's peers. The size of the discrepancy between the target and peer performance levels, and

the tolerance of the significant adults in the child's life are then used in determining the child's eligibility for special services.

How well does performance measurement lend itself to making decisions regarding a child's placement within a curriculum sequence? In the area of reading, for example, Hansen and Lovitt (1976) and Jenkins and Fleischer (1976) have combined performance measurement with informal reading inventory procedures to make student placement decisions. When a student is referred for reading assessment, he or she is asked to read, on three separate days, 100 word samples from each of the books in the reading series. The examiner notes the student's correct words per minute, reading accuracy, and percent of comprehension questions correctly answered. Median performances are calculated for each of these separate measures; then, using some criterion such as 95% reading accuracy, 80% reading comprehension, and 50 words read correctly per minute, a placement decision is made. This reading placement procedure has several advantages. Compared to standardized achievement tests, for example, it is more sensitive to the particular curriculum in which a student will read. In addition, since multiple measures are taken from each book, measurement error is reduced and along with it erroneous placement decisions.

Direct and daily performance measurement has its greatest value in the area of program adjustments. In particular, these lend themselves well to making instructional decisions, incentive decisions, and resource utilization decisions. However, curriculum effectiveness decisions (i.e., is the child making greater progress in one curriculum or another?) are not easily made with such a data system. Such decisions would need to be made on the basis of measurements that are external to

either curriculum. Direct and daily performance measurement, as commonly used by teachers, is not useful in making decisions about the least restrictive environment. To ensure relevance, teachers individualize measurements so highly for their students that the data is curriculum and environment specific. Thus, it is difficult to compare a student's performance in two different environments since changing the curriculum or environment requires changing the measurements.

Daily performance measurement may be used for summative evaluation, but because the measurements are often so individualized by teachers and children, they are often difficult to aggregate. More often than not, children who enter a special education program bring with them no performance data with which to compare the effects of special services. Rather, the special educator is likely to be the first one to institute a direct and daily measurement system. Moreover, norms often are lacking with which to compare a child's performance in any particular area (e.g., addition facts, oral reading, comprehension questions).

Progress or Mastery Monitoring

A second approach to measurement in the least restrictive alternative involves monitoring student mastery of a set of objectives or tasks. In contrast to performance measurement, progress monitoring is designed to describe a student's rate of progress across a series of different tasks over time.

Progress monitoring is well illustrated in the work of Deno and Mirkin (1977). The system is a combination of norm referenced and criterion referenced measurement. It is criterion referenced in the sense that objectives are clearly defined and specific criteria are

set with respect to each of the objectives. For example, an objective might be "adds two one digit numbers, sums to nine, with 95% accuracy," or "reads orally story number four of 'Mustard Seed Magic' with 98% accuracy." This system is also norm referenced in the sense that an individual student's progress (mastery of objectives) is evaluated by comparing it to the progress or rate of mastery of normal peers for the same objectives. An example of this data system is illustrated in Figures 2 and 3.

- -

Insert Figures 2 and 3 about here

- -

The abcissa of the graph in Figure 2 represents the time an individual has been in school, broken down by year and month. The ordinate of the graph represents progress on the objectives in a given curriculum, again broken down by year and month. The diagonal line dividing the graph is meant to show the rate at which a normal student would progress through the curriculum. As can be seen from the graph, the student should make one month's growth in the curriculum for each month in school and, likewise, for each year in school, a student should make one year's growth in the curriculum.

The letters on the ordinate of Figure 2 represent different books in a particular reading series. Inspection of this graph reveals that at the end of the first month of first grade, students should have mastered the contents of book E, and at the end of sixth grade they should have mastered the contents of book L. Figure 3 shows the actual achievement level of a particular student with respect to the achievement of normally progressing peers. The difference between the heavy diagonal line and the student's actual reading level represents the discrepancy in the rate of progress between the student and the student's

peers. This discrepancy can also be numerically described by dividing the normal progress rate by the child's actual progress rate. In the present example, these students have been in school for 18.5 months and should therefore have made 18.5 months progress in the curriculum. The target student shown in this graph, however, has made only four months progress in the curriculum. Dividing 18.5 by four yields a discrepancy ratio of 4.63, which indicates that this child's peers are progressing at a rate 4.63 times faster.

What is needed to employ such a pupil progress data system? First, the essential tasks of the least restrictive environment must be identified. This is not always as easy as it seems, particularly in the case of open education programs which may not employ basal reading series or a uniform math program. Second, a mastery criterion must be established for each of the tasks in the curriculum. Third, an index of normal progress must be obtained. Such an index could be estimated, or, better yet, empirically determined by conducting performance measurement of children who are considered to be making normal progress on the curriculum tasks.

Assuming that all of these conditions have been met, how well does this data system meet the decision needs which we have identified for special education? Determination of eligibility for special services, in terms of achievement, should be quite straightforward with this system, assuming that a school district or state department is willing to identify a discrepancy ratio which students must exceed to qualify for services. For example, students who are progressing on objectives at a rate two times slower than their normally progressing peers might be candidates for special education. As for placement decisions, it

should be a relatively easy matter to place a child within an objectives sequence since a performance criterion for mastery has been identified for all objectives. In our own experience with this system we have used the modified informal reading inventory described earlier as a means for making placement decisions.

A progress monitoring data system can be used in evaluating most types of program adjustments. By inspecting a child's rate of progress, one can evaluate instructional decisions, resource utilization decisions, and least restrictive environment decisions. If progress monitoring is designed to describe rate of mastery in a specific curriculum, this system is not useful for evaluating curriculum adjustments. Since in those cases, the child's progress is measured in relation to a particular curriculum sequence, it is impossible to compare the effects of different curricula. If progress monitoring is based on objectives apart from any specific curriculum, however, the influence of curriculum adjustments on the rate at which those objectives are mastered can easily be evaluated.

At first glance, it would seem that overall program evaluation could not be conducted using the progress monitoring system. If different districts, schools, and teachers were using different sets of objectives, it would seem that there would be no basis for comparability. This problem may be more apparent than real, however, since we believe that it may be possible to use discrepancy rates as the basic datum in overall program evaluation. Even though different objectives or curricula are used in different schools, every child's progress could be compared to the progress of his or her normally progressing peers,

who, by definition, are making one month's progress in a curriculum for every month in school. Thus, it would appear that program evaluators could assess the impacts of particular programs or combinations of programs by determining the extent to which these programs reduced the disciplinary ratios of handicapped students.

In terms of the characteristics of an ideal data system, the progress monitoring system has much to offer. By definition, the system measures exactly what is taught. For that same reason, the system should be quite sensitive to changes in rate of progress. The system is flexible in that both the objectives and the performance levels considered necessary for mastery can be defined by the significant adults in a child's life. It is easy to conceive of instances where curriculum tasks and levels of performances on these tasks would vary from school to school, depending upon the biases and values held by particular adults in those schools. While this at first might seem a bit artibrary, it is probably important to reflect the situation specific variables that affect children's assignment to special education. Data in this system can be gathered as frequently as necessary -- daily, weekly, monthly, and annually. Moreover, children can be assessed at any particular time and normative data for that assessment period can be generated by the school. In such a case, it makes no difference whether a child enters the system in September or November or February or May, since normal progress can be determined at any point in time.

The validity of the progress monitoring data system depends to a large extent on the school's program. If progress on the school's objectives reflects progress towards adult functioning, then this system

is validated. If, however, progress on the objectives is not positively and highly correlated with progress toward adult functioning, then it makes no sense to judge a program's worth on the basis of progress rate. Whether the school objectives are valid should be an issue in the development of the IEP and its acceptance by the parents and the teaching professionals responsible for a given child.

Conclusion

In this paper we have considered the reasons for increased emphasis on measuring pupil progress, the uses made of pupil progress data, the desirable characteristics of measurement systems, and the problems associated with currently commercially available data systems. In addition, we have identified techniques for measurement based on the requirements of the least restrictive environment. We believe that measurement rooted in the least restrictive environment is not only more consistent with the intents of P.L. 94-142 and more useful to special educators, but also that such measurement better meets the requirements for technical adequacy than most standardized tests. At the same time, we recognize that data systems tailored by teachers to the requirements of the least restrictive environment must pass the test of consumer usefulness, and we are sure that as they are tested, they will be improved. Our concern now is that such testing and improvement should move ahead more rapidly than has been the case.

References

Armbruster, B. B., Stevens, R. J., & Rosenshine, B. Analyzing content coverage and emphasis: A study of three curricula and two tests. Technical Report No. 26. Center for the Study of Reading, University of Illinois, Urbana-Champaign, 1977.

Deno, S. L., & Mirkin, P. Data based program modification. Reston, Va.: Council for Exceptional Children, 1977.

Eaton, M., & Lovitt, T. Achievement tests vs. direct and daily measurement. In G. Semb (Ed.) Behavior analysis and education. Lawrence, Kans.: University of Kansas, 1972.

Goldstein, H., Moss, J., & Jordan, L. J. The efficacy of special class training on the development of mentally retarded children. Cooperative Research Project No. 619. Washington, D.C.: U.S. Office of Education, 1965.

Goodlad, J. I., & Klein, M. F. Looking behind the classroom door. Worthington, Ohio: Charles A. Jones Publishing Co., 1974.

Hansen, C., & Lovitt, T. Round one - placing the child in the right reader. Journal of Learning Disabilities, 1976, 9, 347-353.

Institute for Research on Teaching. Experts' diagnoses inconsistent. Communication Quarterly, 2(3), Spring 1979.

Jenkins, J., & Fleischer, L. Validation studies of curriculum based assessment for reading. Unpublished paper. Experimental Education Unit, University of Washington, 1976.

Jenkins, J., & Pany, D. Standardized achievement tests: How useful for special education? Exceptional Children, 1978, 44(6), 448-453.

Mirkin, P. K. A comparison of the effects of three formative evaluation strategies on reading performance. Doctoral dissertation, University of Minnesota, 1978.

Peterson, P. L., Marx, C. W., & Clark, R. M. Teacher planning, teacher behavior, and student achievement. American Educational Research Journal, 1978, 15(3), 417-432.

Popham, W. J., & Baker, E. Establishing instructional goals. Englewood Cliffs, N. J.: Prentice Hall, 1970.

Tallmadge, G. K. Interpreting NCE's. Technical Paper #2 ESEA Title 1 Evaluation and Monitoring System. Mountain View, Ca.: RMC Research Corporation, 1976.

Tallmadge, G. K. Cautions to evaluators. In M. Wargo and D. Green (Eds.) Achievement testing of disadvantaged and minority students for educational program evaluation. McGraw-Hill, 1977.

White, O. R., & Haring, N. G. Exceptional teaching: A multimedia training package. Columbus, Ohio: Charles E. Merrill, 1976.

Zoharik, J. A. Teachers planning models. AERA, Washington, D. C., April 1975.

Table 1

Grade Equivalent Scores Determined by Matching Specific Reading Text Words to Standardized Reading Test Words

Curriculum	Test						
	PIAT	MAT		SORT	SAT		WRAT
		Word Knowledge	Word Analysis		Word Reading	Paragraph Meaning	
Bank Street Readers							
Grade 1	1.5	1.0	1.1	1.8	1.0	1.6	2.0
Grade 2	2.8	2.5	1.2	2.9	-	2.9	2.7
Economy*							
Grade 1	2.0 (3.3)	1.4 (1.9)	1.2 (1.8)	2.2 (2.7)	1.5 (1.7)	1.8 (2.3)	2.2 (3.0)
Grade 2	3.3 (3.8)	1.9 (3.7)	< 1.0 (2.0)	3.0 (3.5)	- -	2.6 (3.2)	3.0 (3.6)
Ginn 360**							
Grade 1	1.5 (1.5)	< 1.0 (< 1.0)	< 1.0 (1.0)	1.4 (1.5)	< 1.0 (< 1.0)	1.5 (1.5)	1.7 (1.8)
Grade 2	2.2 (2.8)	2.1 (2.5)	< 1.0 (1.1)	2.7 (2.7)	- -	1.9 (1.9)	2.3 (2.5)
Houghton-Mifflin***							
Grade 1	2.0 (3.1)	1.1 (1.9)	< 1.0 (1.7)	1.6 (2.9)	< 1.0 (1.7)	1.6 (2.4)	2.0 (3.1)
Grade 2	3.4 (3.8)	2.3 (2.9)	< 1.0 (2.0)	2.4 (3.4)	- -	2.3 (3.4)	2.4 (2.9)
Macmillan							
Grade 1	1.8	1.1	1.0	1.9	1.1	1.6	2.3
Grade 2	2.2	2.5	1.1	2.9	-	2.5	2.6
SRA							
Grade 1	1.5	1.2	1.3	1.0	1.0	1.6	2.1
Grade 2	3.1	2.5	1.4	2.9	-	2.9	3.5
Sullivan							
Grade 1	1.8	1.4	1.2	1.1	1.2	1.6	2.0
Grade 2	2.2	2.4	1.1	2.5	-	3.1	2.5
Grade 3	3.8	-	-	4.3	-	-	3.5

*Scores in parentheses reflect the inclusion of words listed as "sounding vocabulary."
** " " " " " " " " " " "enrichment."
*** " " " " " " " " " " "supplementary."

Note. From "Standardized achievement tests: How useful for special education?" by J. Jenkins and D. Pany, Exceptional Children, 1978, 44(6), 448-453.

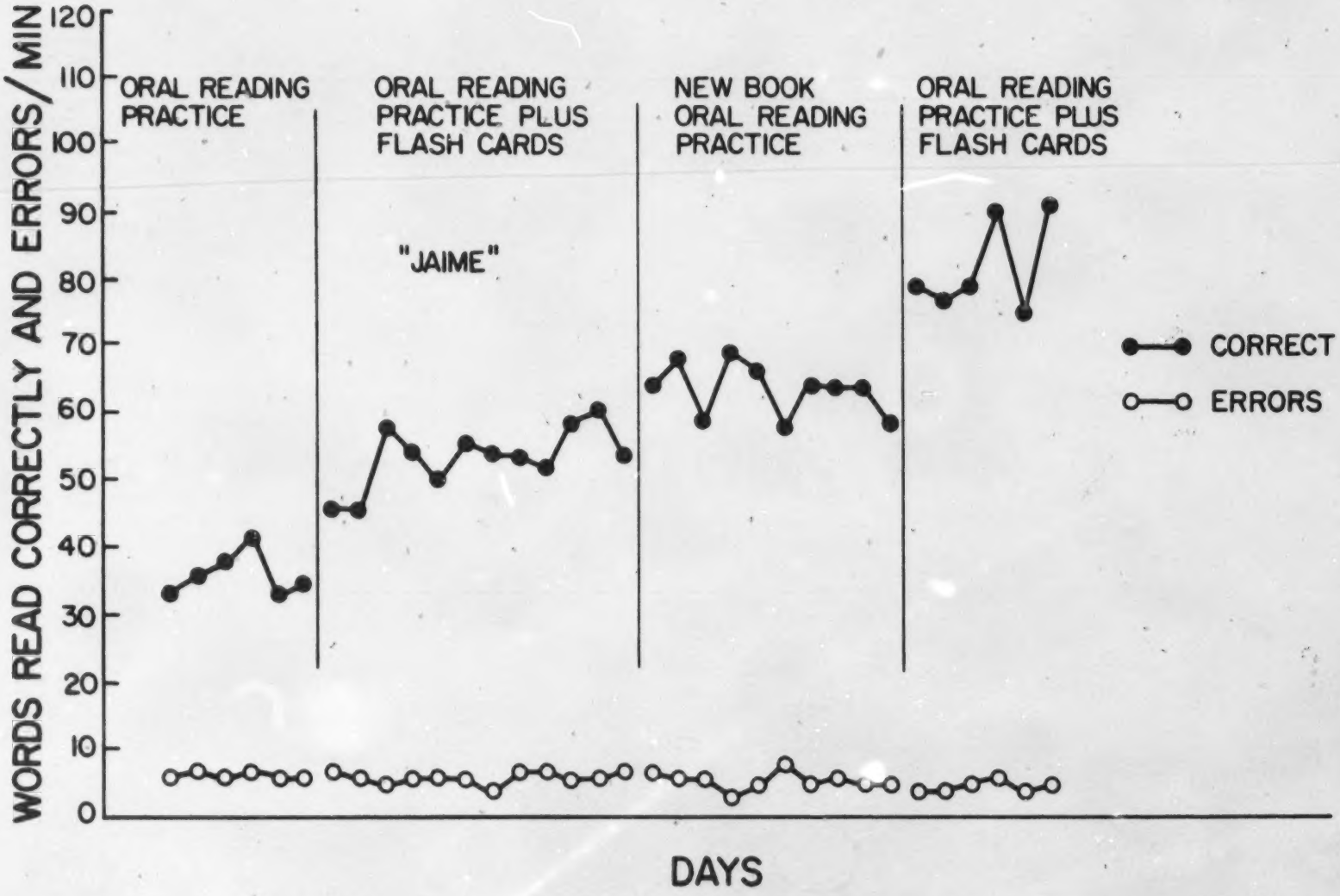
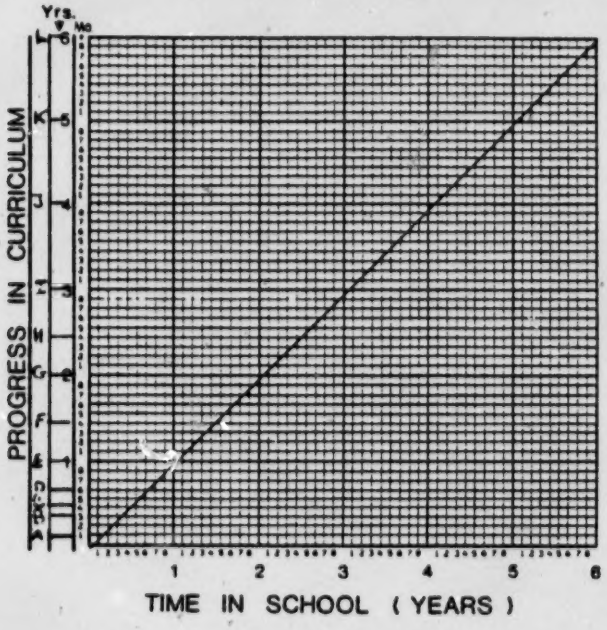


Figure 1. Oral reading correct and error performance for Jaime. Vertical lines indicate a change in the instructional program.

Figure 2

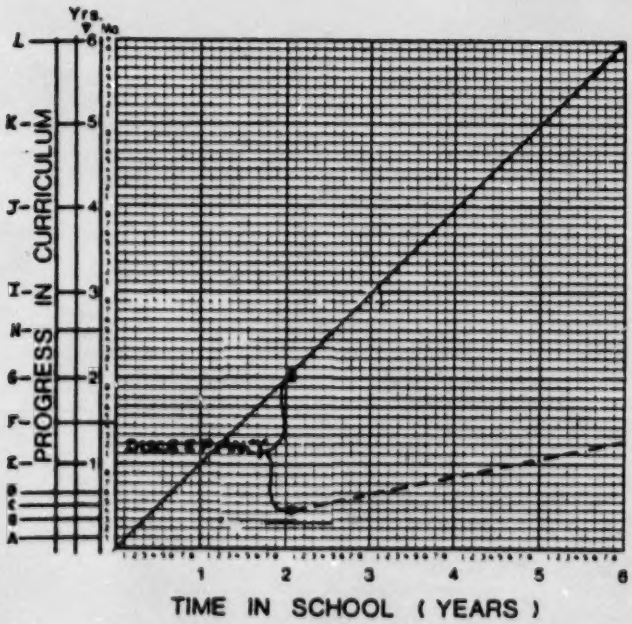
STUDENT: Average SCHOOL: River Run CURRICULUM Read Series



Six-year graph showing expected progress for average students for Read Series (American Book Co., 1968), River Run School.

Figure 3

STUDENT: RICKY SCHOOL: RIVER RUN CURRICULUM: READ SERIES



KEY:
Δ DESIRED LEVEL of MASTERY for AGE/GRADE
O ACTUAL LEVEL of MASTERY
/ DESIRED PROGRESS LINE
/ NONINTERVENTION PROGRESS LINE

Actual mastery level, estimated nonintervention progress line, and desired progress for target student in Read Series (American Book Co., 1968), River Run School.

PUBLICATIONS

Institute for Research on Learning Disabilities
University of Minnesota

The Institute is not funded for the distribution of its publications. Publications may be obtained for $2 per document, a fee designed to cover printing and postage costs. Only checks and money orders payable to the University of Minnesota can be accepted. All orders must be prepaid.

Requests should be directed to: Editor
Institute for Research on Learning Disabilities
350 Elliott Hall
75 East River Road
University of Minnesota
Minneapolis, Minnesota 55455

Ysseldyke, J. E. Assessing the learning disabled youngster: The state of the art (Research Report No. 1). November, 1977.

*Ysseldyke, J. E. Research and development plan: University of Minnesota Institute for Research on Learning Disabilities (Monograph No. 1). June, 1978.

*Ysseldyke, J. E., & Thurlow, M. L. (Eds.). Synthesis of the knowledge base: Identification and assessment of learning disabled children (Monograph No. 2). June, 1978.

*Ysseldyke, J. E. Organizational plan for Minnesota's Institute for Research on Learning Disabilities (Monograph No. 3). June, 1978.

*Ysseldyke, J. E., & Thurlow, M. L. Specific investigations to be completed during years two and three (Monograph No. 4). June, 1978.

*Ysseldyke, J. E., & Thurlow, M. L. Training opportunities at Minnesota's Institute for Research on Learning Disabilities (Monograph No. 5). June, 1978.

*Ysseldyke, J. E. Client base (Monograph No. 6). June, 1978.

Ysseldyke, J. E., & Regan, R. R. Nondiscriminatory assessment and decision making (Monograph No. 7). February, 1979.

*As part of its continuation proposal, the Institute was required to prepare these monographs. Because they are part of the proposal, they are not available for general distribution.

**Thurlow, M., & Ysseldyke, J. Assessment in the child service demonstration centers for learning disabled children (Research Report No. 2). March, 1979.

Foster, G., Algozzine, B., & Ysseldyke, J. Susceptibility to stereotypic bias (Research Report No. 3). March, 1979.

Algozzine, B. An analysis of the disturbingness and acceptability of behaviors as a function of diagnostic label (Research Report No. 4). March, 1979.

Algozzine, B., & McGraw, K. Diagnostic testing in mathematics: An extension of the PIAT? (Research Report No. 5). March, 1979.

Deno, S. L. A direct observation approach to measuring classroom behavior: Procedures and application (Research Report No. 6). April, 1979.

Ysseldyke, J. E., & Mirkin, P. K. Proceedings of the Minnesota roundtable conference on assessment of learning disabled children (Monograph No. 8). April, 1979.

Somwaru, J. P. A new approach to the assessment of learning disabilities (Monograph No. 9). April, 1979.

Algozzine, B., Forgnone, C., Mercer, C. D., & Trifiletti, J. J. Toward defining discrepancies for specific learning disabilities: An analysis and alternatives (Research Report No. 7). June, 1979.

Algozzine, B. The disturbing child: A validation report (Research Report No. 8). June, 1979.

Ysseldyke, J. E., Algozzine, B., Regan, R., & Potter, M. Technical adequacy of tests used by professionals in simulated decision making (Research Report No. 9). July, 1979.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. Measuring pupil progress toward the least restrictive environment (Monograph No. 10). August, 1979.

**This research report is not being distributed by the Institute. Requests for it should be directed to: The NETWORK, 290 S. Main Street, Andover, Massachusetts 01810.